

So ziehen Sie größeren Nutzen aus Ihren Umfragedaten

Entdecken Sie vier fortschrittliche Analysetechniken, die Umfrageforschung noch effektiver machen



Inhalt:

- 1 Einführung
 - 2 Deskriptive Umfrageforschung
 - 3 Segmentieren von Umfrageteilnehmern mit Cluster-Analyse
 - 6 Gruppieren von Fragen mit der Faktoranalyse
 - 7 Ermittlung der geeigneten Analysetechnik mit IBM SPSS Statistics
 - 12 Vorhersagen mit Regression
 - 15 Zusammenfassung
 - 16 Informationen zu IBM Business Analytics
-

Einführung

Wie bei vielen Umfrageforschern gehören sicher auch bei Ihnen Kreuztabellen, Balkendiagramme und die Ermittlung der Mittelwertdifferenzen zwischen Gruppen zu den täglich verwendeten Tools und Techniken für die Datenanalyse. Doch so wertvoll diese Methoden auch sind, so sind sie möglicherweise doch zu vereinfachend, sodass Sie nicht den größtmöglichen Nutzen aus Ihren Umfragedaten ziehen können.

In diesem Whitepaper werden vier Typen von fortschrittlichen Analysen vorgestellt – Cluster, Faktor, Reliabilität und Regression – mit denen Sie wichtige Einblicke gewinnen können, die Ihnen mit grundlegenderen Methoden möglicherweise entgehen. Wenn Sie Ihre Tools für Umfrageanalysen erweitern, können Sie tiefer in die Daten vordringen und Umfrageteilnehmer und -teilnehmer noch besser verstehen, bessere Maßzahlen für wichtige Konzepte erstellen und genauere Vorhersagen über Verhaltensweisen und Haltungen treffen.

Im Folgenden finden Sie einen kurzen Überblick über die vier Techniken:

- Cluster-Analyse wird verwendet, um ähnliche Gruppen bzw. Segmente von Umfrageteilnehmern zu ermitteln. Dank der Segmentierung können Sie Ihre Vertriebs- und Marketingbestrebungen gezielt auf die definierten Gruppen richten. Außerdem können Sie Untergruppen in Analysen verwenden, um eine höhere Sensitivität für die Unterschiede zwischen den Umfrageteilnehmern zu erreichen.
- Mit der Faktor- und der Reliabilitätsanalyse können Sie mehrere Fragen kombinieren und so eine validere und zuverlässigere Maßzahl für ein wichtiges Konzept erhalten. Außerdem können Sie damit leichter Fragen isolieren, die redundant oder überflüssig sind.
- Mit der Regressionsanalyse können Sie Vorhersagemodelle für Verhaltensweisen erstellen, die viele Prädiktorvariablen gleichzeitig enthalten können. Die Regressionsanalyse ermöglicht die Ermittlung der besten Prädiktoren, auf die Sie sich dann in zukünftigen Aktionen konzentrieren können.



Highlights:

Cluster-, Faktor-, Reliabilitäts- und Regressionsanalyse sind fortschrittliche Analysemethoden, mit denen Sie wichtige Einblicke gewinnen können, die Ihnen mit grundlegenden Methoden möglicherweise entgehen.

Wenn diese Techniken für Sie neu sind, kann es hilfreich sein, sich beim Lesen dieses Dokuments auf die Vorteile der einzelnen Methoden zu konzentrieren und weniger auf die technischen Einzelheiten zu achten. Wenn Sie diese fortschrittlichen Techniken in Ihren eigenen Analysen ausprobieren möchten, sollten Sie mit der Methode beginnen, die sich am besten für Ihre Daten eignet oder die Ihnen am leichtesten fällt.

IBM SPSS-Software und -Technologien werden im gesamten Dokument verwendet, um die Anwendung fortschrittlicher Analysemethoden auf typische Umfragedaten zu illustrieren. Die einzelnen Beispiele enthalten jeweils auch Ratschläge zur Verwendung der Technik sowie zur Interpretation der Ergebnisse und Ausgaben.

Deskriptive Umfrageforschung

Alle Umfrageforscher verwenden deskriptive statistische Methoden, um Daten zusammenzufassen und eine Beschreibung der Antworten auf Fragen zu erhalten. Zu diesen Methoden gehören Häufigkeitstabellen, Kreuztabellen (Tabellen mit Spalten und Zeilen) und die Ermittlung der Mittelwertdifferenzen zwischen Gruppen bzw. Korrelationen zwischen Fragen. Bei Abbildung 1 beispielsweise handelt es sich um eine typische Häufigkeitstabelle, die die Gesamtzufriedenheit der Kunden für die hypothetische Hotelkette BSI zeigt.

I am a satisfied BSI customer

	Frequency	Percent	Valid Percent	Cumulative Percent
Strongly Disagree	20	2.1	2.1	2.1
Disagree	130	13.6	13.6	15.7
Neither Agree Nor Disagree	360	37.7	37.7	53.4
Agree	300	31.4	31.4	84.8
Strongly Agree	145	15.2	15.2	100.0
Total	955	100.0	100.0	

Abbildung 1: Diese Standardtabelle zur Häufigkeit zeigt Haltungen in Bezug auf Gesamtzufriedenheit der Kunden.

Diese Art von Analyse ist zwar bei jedem Umfrageprojekt notwendig, doch eine Reihe solcher Tabellen, oder auch Kreuztabellen, kann nur begrenzte Informationen über die Haltungen und Verhaltensweisen der Umfrageteilnehmer bieten. Dies liegt daran, dass die reale Welt multivariat ist, dass also bei einer einzelnen Antwort jeweils viele Faktoren eine Rolle spielen. Bivariate Ansätze, wie beispielsweise die Analyse der Daten Frage für Frage oder die Verwendung einer Kreuztabelle zur Ermittlung, ob zwei Fragen in Bezug zueinander stehen, führen zu einer zu sehr vereinfachten Sicht auf den Kunden.

Mit der Cluster-Analyse können Sie Umfrageteilnehmer anhand verschiedener Merkmale zu Gruppen zusammenfassen, um bessere Einblicke in wichtige Ähnlichkeiten und Unterschiede zu erhalten.

Die Tabelle in Abbildung 1 enthält keine Informationen darüber, welche Kunden mehrere Fragen ähnlich beantworteten. Wenn nur eine oder zwei Fragen analysiert werden, ist es nicht möglich zu ermitteln, welche Fragengruppen ähnliche Konzepte messen. Wenn eine Tabelle verwendet wird, um zu ermitteln, inwieweit sich anhand von Antworten auf eine Frage die Antworten auf eine zweite Frage vorhersagen lassen, werden außerdem die Faktoren ignoriert, die die zweite Frage beeinflussen. Wenn also beispielsweise die Bewertung des allgemeinen Service verwendet wird, um die Zufriedenheit mit der Hotelkette vorherzusagen, werden andere Faktoren, wie die Häufigkeit der Übernachtung, die Bewertung der Restaurants und die Bewertung der Zimmerqualität ignoriert.

Mit den in diesem Dokument erörterten fortschrittlichen Umfragemethoden können Sie viele Fragen gleichzeitig analysieren, um die Umfrageteilnehmer zu Clustern zusammenzufassen, Fragen zu gruppieren und Vorhersagen mit größerer Genauigkeit zu treffen.

Segmentieren von Umfrageteilnehmern mit Cluster-Analyse

Mit der Cluster-Analyse können Sie Umfrageteilnehmer mit ähnlichen Verhaltensweisen, Präferenzen oder Merkmalen zu Clustern bzw. Segmenten zusammenzufassen. Durch Segmentierung erhalten Sie bessere Einblicke in wichtige Ähnlichkeiten und Unterschiede zwischen den Umfrageteilnehmern. Mit diesen Informationen können Sie gezielte Marketingstrategien entwickeln oder Untergruppen für die Analyse erstellen. Bei Umfragedaten können die Forscher durch Clusterbildung Umfrageteilnehmer zu Gruppen zusammenfassen, die auf mehrere Fragen ähnliche Antworten abgegeben haben.

Die Clusterbildung bzw. Segmentierung ist eine multivariate Technik, bei der Antworten auf mehrere Fragen analysiert werden, um ähnliche Umfrageteilnehmer zu finden. Clusterbildung beruht auf dem Konzept der Gruppenerstellung anhand der Nähe bzw. Distanz untereinander. Umfrageteilnehmer innerhalb eines Clusters sind daher relativ homogen.

Es gibt zwei Typen von Cluster-Analysen:

- Hierarchisch: Beobachtungen werden in einem Cluster verbunden und bleiben dies während der gesamten Clusterbildung.
- Nichthierarchisch: Die Fälle können beim Fortschreiten der Clusterbildung den Cluster wechseln. Die am häufigsten verwendete nichthierarchische Methode ist K-Means.

Für die Clusteranalyse müssen folgende Aufgaben ausgeführt werden:

- Überprüfung der Anzahl der Umfrageteilnehmer in jedem Cluster, da Cluster mit nur sehr wenigen Umfrageteilnehmern nicht sonderlich nützlich sind
- Bewertung, ob die Cluster sinnvoll sind und ob ihre Merkmale leicht zu verstehen und zu beschreiben sind
- Validieren der Cluster durch Analysieren ihres Bezugs zu anderen Variablen

Beim folgenden Beispiel für die Cluster-Analyse wird das Two-Step-Cluster-Verfahren aus IBM SPSS Statistics verwendet, bei dem statistische Kriterien einbezogen werden, um die optimale Cluster-Anzahl zu ermitteln.

In diesem Beispiel wenden wir die Two-Step-Cluster-Methode auf Kundenumfragedaten für die hypothetische Hotelkette BSI an. Die Messungen für die Fragen erfolgte zwar anhand verschiedener Skalen, dies stellt jedoch kein Problem dar, da für die Two-Step-Cluster-Methode nominal-, ordinal- und intervallskalierte Daten verwendet werden können. Wir haben vor, Cluster von Kunden anhand der folgenden Kriterien zu erstellen:

- Häufigkeit der Übernachtungen in BSI-Hotels
- Dauer der Kundenbeziehung
- Brauchbarkeit des Internetzugangs
- Unternehmenstyp des Kunden
- Grad der Einbindung in die Entscheidung des Unternehmens für BSI-Hotels
- Wichtigkeit der Reisen für die Tätigkeit

Die erste Ausgabe aus der Two-Step-Methode zeigt, wie viele Cluster gefunden wurden und wie viele Umfrageteilnehmer sich in den einzelnen Clustern befinden. Aus der Tabelle in Abbildung 2 können wir entnehmen, dass mit der Two-Step-Methode drei natürliche Gruppen oder Cluster in den Daten gefunden wurden. Diese beruhen auf den Antworten auf die sechs oben genannten Fragen.

Cluster Distribution

	N	% of Combined	% of Total
Cluster 1	180	26.9%	18.8%
2	240	35.8%	25.1%
3	250	37.3%	26.2%
Combined	670	100.0%	70.2%
Excluded Cases	285		29.8%
Total	955		100.0%

Abbildung 2: Mit der Two-Step-Cluster-Methode wurden drei Kundencluster identifiziert, die jeweils ungefähr dieselbe Anzahl von Kunden enthalten.

Anhand von weiteren Informationen aus der Two-Step-Methode können wir verstehen, für welchen Kundentyp die einzelnen Cluster jeweils stehen. In Abbildung 3 auf der nächsten Seite sehen wir beispielsweise, dass die Kunden in Cluster 1 häufiger beruflich reisen als die Kunden in den beiden anderen Clustern und dass sie häufig in BSI-Hotels übernachten. Gemessen an der Dauer der Kundenbeziehung jedoch sind die Kunden in Cluster 1 nicht die treuesten.

	How long have you been staying at BSI hotels?	Travel an integral prt of wrk	How frequently do stay at BSI hotels?	Usefulness of internet services
Cluster	Mean	Mean	Mean	Mean
1	3.44	4.36	5.33	3.33
2	3.00	3.69	4.23	3.52
3	3.82	4.08	4.50	3.66
Combined	3.43	4.01	4.63	3.52

Abbildung 3: In dieser Tabelle wird der Mittelwert für verschiedene Fragen nach Cluster geordnet angezeigt. Es zeigt sich, dass die Kunden in Cluster 1 häufiger in BSI-Hotels übernachten und häufiger beruflich reisen als die Kunden in den beiden anderen Clustern.

Wenn Sie der Ansicht sind, die Cluster und ihre Merkmale hinreichend verstanden zu haben, können Sie aussagekräftige Beschriftungen erstellen. In diesem Fall beschriften wir die Kunden in Cluster 1 als „Frequent BSI Road Warriors“ (Regelmäßig bei BSI übernachtende Geschäftsreisende). Die Kunden in Cluster 2 sind „Less Frequent Travelers“ (Weniger häufig Reisende) und die Kunden in Cluster 3 sind „Loyal BSI Road Warriors“ (BSI-treue Geschäftsreisende).

Verwendung der Cluster in der Analyse

Nachdem wir nun die Cluster der Umfrageteilnehmer ermittelt haben, können wir sie in Analysen und Berichten verwenden. Um beispielsweise zu ermitteln, in welcher Relation die Gesamtzufriedenheit der Kunden mit der Cluster-Zugehörigkeit steht, haben wir ein gruppiertes Balkendiagramm erstellt (Abbildung 4). Das Diagramm zeigt, dass die Kunden in Cluster 3 zufriedener mit BSI-Hotels sind als die Kunden in Cluster 1. Da die Kunden in Cluster 1 häufiger in BSI-Hotels übernachten als die Kunden in den anderen Clustern, können diese Ergebnisse auf einen Handlungsbedarf bei diesem Segment hindeuten. Die Merkmale der zufriedensten und am wenigsten zufriedenen Kunden zu kennen, kann wichtige Unternehmensentscheidungen erleichtern.

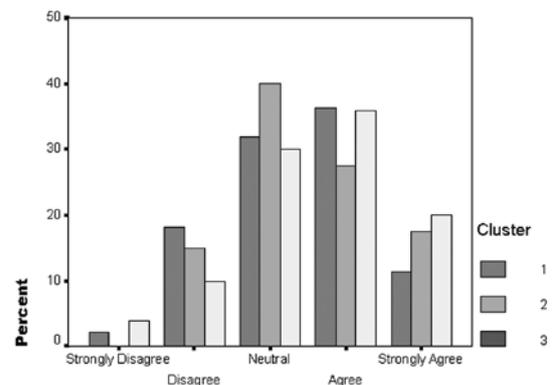


Abbildung 4: Dieses Balkendiagramm zeigt die Unterschiede in der Gesamtzufriedenheit zwischen den drei Kundenclustern. Die Kunden in Cluster 3 sind am zufriedensten, während die Kunden in Cluster 1 am unzufriedensten sind.

Präsentation der Ergebnisse

Mit der Prozedur „Benutzerdefinierte Tabellen“ in IBM SPSS Statistics können Sie ansprechende und komplexe Tabellen mit Spalten und Zeilen erstellen. Die Tabelle in Abbildung 5 zeigt den Mittelwert der Kundenzufriedenheit in Abhängigkeit von der Cluster-Zugehörigkeit. Für jeden Cluster wird die Bereitschaft des Kunden, in einem BSI-Hotel zu übernachten, verschachtelt. Die einzelnen Zeilen enthalten Fragen zur Nutzung von Hotel-Spas und der Lösung bei Problemen mit der Rechnungsstellung. Mit der Prozedur „Benutzerdefinierte Tabellen“ können Sie viele verschiedenartige Informationen in einer ansprechenden und kompakten Tabelle zusammenfassen.

			TwoStep Cluster					
			Frequent BSI Road Warriors		Less Frequent Travelers		Loyal BSI Road Warriors	
			Would you stay with BSI again?		Would you stay with BSI again?		Would you stay with BSI again?	
			Yes	No	Yes	No	Yes	No
I am a satisfied BSI customer	Use spa facilities?	No	3.9	2.0	3.8	2.0	4.0	2.5
		Yes	3.6	.	3.6	2.0	3.6	1.0
	Ever called BSI to resolve a problem?	Yes	3.8	2.0	3.9	2.0	3.6	2.0
		No	3.6	2.0	3.6	2.0	4.2	1.7

Abbildung 5: Mit der Prozedur „Benutzerdefinierte Tabellen“ können Sie komplexe Tabellen, wie die oben abgebildete, erstellen. Diese Tabelle verdeutlicht die Beziehung zwischen den Kunden in den einzelnen Clustern und der Gesamtzufriedenheit und schließt mehrere verschachtelte Fragen mit ein.

Zusammengesetzte Maßzahlen für kritische Konzepte können zu aussagekräftigeren Analysen führen. Die Faktoranalyse bietet die Möglichkeit, Cluster bzw. Gruppen von Fragen zu ähnlichen Konzepten zu ermitteln.

Gruppieren von Fragen mit der Faktoranalyse

Die meisten Fragebogen enthalten zu jedem Hauptpunkt mehrere Fragen. Ein Fragebogen zur Messung der Patientenzufriedenheit mit einem Krankenhaus beispielsweise könnte mehrere Fragen zur ärztlichen Sorgfalt sowie mehrere Fragen zur Leistung des sonstigen Personals, wie beispielsweise der Radiologieassistenten, enthalten. So sinnvoll die Einzelbetrachtung jeder Frage auch sein mag, es lassen sich doch häufig zuverlässigere und validere Maßzahlen erzielen, wenn die Antworten auf mehrere Fragen gleichzeitig verwendet werden. Zusammengesetzte Maßzahlen für kritische Konzepte können aussagekräftigere Analysen ergeben.

Die Faktoranalyse bietet die Möglichkeit, Cluster bzw. Gruppen von Fragen zu ähnlichen Konzepten zu ermitteln, die auf Korrelationen oder Kovarianzen zwischen einzelnen Fragen beruhen. Sie können die Faktoren für folgende Aufgaben nutzen:

- Erstellen von Skalen bzw. zusammengesetzten Maßzahlen aus mehreren Fragen.
- Verringern der Fragenzahl in einem Fragebogen, beispielsweise durch Ermittlung von Fragen, die dasselbe Konzept messen.
- Gewinnen von Einblicken in die Beziehungen zwischen mehreren Fragen gleichzeitig

Die Faktoranalyse ist eine Technik vom Typ „Allgemeines lineares Modell“. Dies bedeutet, dass davon ausgegangen wird, dass die Daten auf einer Intervallskala gemessen wurden. Wie in der Umfrageforschung üblich, können problemlos Variablen auf Skalen mit fünf, sechs oder sieben Punkten verwendet werden.

Ermittlung der geeigneten Analysetechnik mit IBM SPSS Statistics

Mit dem Statistik-Assistenten wird die Auswahl der geeignetsten Methode für Ihre Daten zum Kinderspiel. In diesem Fall möchten wir Fragen gruppieren, wir sind jedoch noch nicht sicher, welche Methode verwendet werden soll. Im Menü „Help“ (Hilfe) wählen wir die Option „Statistics Coach“ (Statistik-Assistent), um den ersten Bildschirm aus Abbildung 6 anzuzeigen. Anschließend wählen wir die Option „Identify groups of similar variables“ (Gruppen von ähnlichen Variablen identifizieren), klicken auf „Next“ (Weiter) und wählen den Datentyp aus. Das Dialogfeld „Factor Analysis“ (Faktoranalyse) wird geöffnet, wie im zweiten Bildschirm zu sehen.

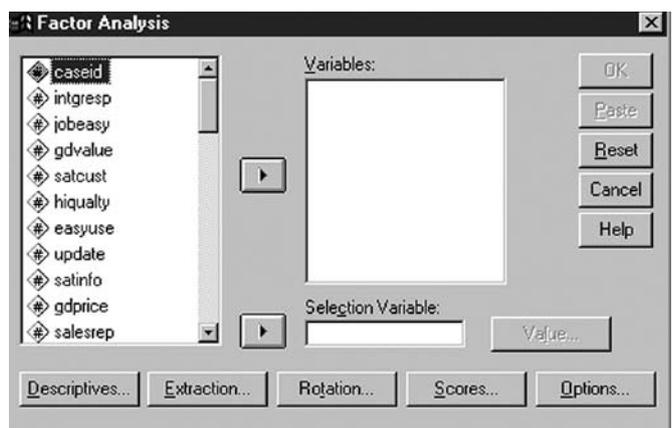
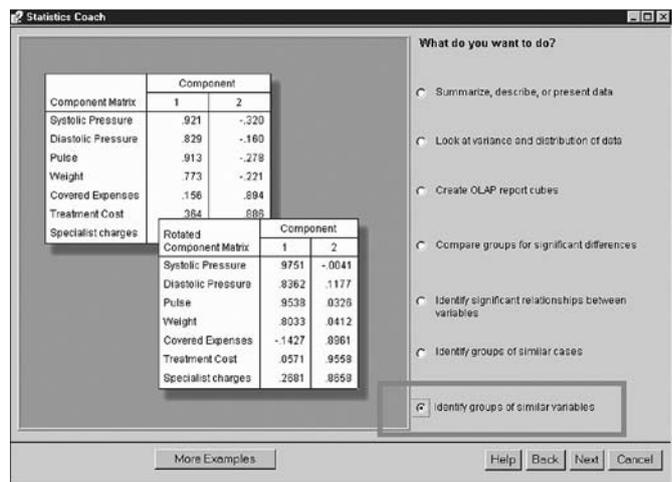


Abbildung 6: Der obere Screenshot zeigt das Dialogfeld „Statistics Coach“ (Statistik-Assistent), in dem die Option „Identify groups of similar variables“ (Gruppen von ähnlichen Variablen identifizieren) ausgewählt ist. Nach der Eingabe des Datentyps wird automatisch das unten abgebildete Dialogfeld „Factor Analysis“ (Faktoranalyse) geöffnet.

Zur Faktoranalyse gehören zwei getrennte Schritte. Im ersten Schritt wird eine kleine Anzahl von Faktoren aus den Daten extrahiert. Stellen Sie sich die Faktoren als die zugrunde liegenden Haltungen vor, die sich in den Antworten auf bestimmten Fragen niederschlagen. Es gibt mehrere Extraktionsmethoden. Die am häufigsten verwendeten sind die Hauptkomponentenextraktion und die Hauptachsen-Faktorenanalyse. Im zweiten Schritt, werden die Faktoren rotiert, um die Interpretation zu erleichtern. Die bei Umfragedaten am häufigsten verwendete Rotationsmethode ist Varimax.

In diesem Beispiel wenden wir die Hauptkomponentenextraktion und die Varimax-Rotation auf die folgenden Aussagen aus der Umfrage unter den Kunden der BSI-Hotelkette an. Für diese Aussagen gab es eine Antwortskala mit fünf Punkten, von „Stimme ganz und gar nicht zu“ bis „Stimme voll und ganz zu“:

- BSI bietet bei seinen Services ein gutes Preis-Leistungsverhältnis
- BSI bietet hochwertige Serviceleistungen
- Es ist einfach, bei BSI Reservierungen vorzunehmen
- BSI erleichtert meine Arbeit
- Ich bin zufrieden mit den BSI-Gastronomieeinrichtungen
- Die Einrichtungen bei BSI sind auf dem neusten Stand
- Der Preis der Zimmer bei BSI ist angemessen

Die Resultierende Tabelle (Abbildung 7) enthält die beiden aus den Umfrageaussagen extrahierten Faktoren. Die Prozedur ermittelt automatisch Faktoren, die einen größeren Teil der Varianz erklären als Einzelaussagen. Die beiden Faktoren in der Tabelle erklären ca. 59 Prozent der Gesamtvarianz zwischen den Aussagen, was ein recht zufriedenstellender Wert ist.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.047	43.529	43.529	3.047	43.529	43.529	2.516	35.941	35.941
2	1.093	15.614	59.143	1.093	15.614	59.143	1.624	23.202	59.143
3	.869	12.268	71.411						
4	.670	9.268	80.678						
5	.544	7.772	88.451						
6	.463	6.608	95.059						
7	.325	4.641	100.000						

Extraction Method: Principal Component Analysis.

Abbildung 7: In dieser Tabelle für die Gesamtvarianz sind die Anzahl der ausgewählten Faktoren und die zugehörige Varianz, jeweils vor und nach der Rotation, abgegeben.

Die wichtigste Ausgabe für unsere Faktoranalyse ist jedoch die rotierte Komponententabelle (Abbildung 8), die die Ladung bzw. Korrelation zwischen den einzelnen Fragen und den beiden extrahierten Faktoren zeigt. Fragen mit hohen Ladungen bei einem Faktor und niedrigen Ladungen bei anderen Faktoren werden dem Faktor mit hoher Ladung zugeordnet. Die Aussage „BSI makes it easy to make reservations“ (Es ist einfach, bei BSI Reservierungen vorzunehmen) ist beispielsweise Faktor 1 zugeordnet, da sie eine hohe Korrelation (0,716) mit diesem Faktor aufweist und eine niedrige Korrelation (0,260) mit Faktor 2.

Aus der Tabelle in Abbildung 8 ist ersichtlich, dass Faktor 1 den ersten fünf Aussagen zugeordnet ist. Faktor 2 ist deutlich den beiden Aussagen zu Zimmerpreis und Preis-Leistungs-Verhältnis zugeordnet.

	Component	
	1	2
BSI makes it easy to make reservations	.716	.260
BSI offers high-quality services	.711	.306
Business Suites makes job easier	.703	.198
BSI facilities are kept up to date	.626	-.413
Satisfied with dining facilities?	.596	.194
BSI rooms are priced right	.148	.842
BSI services are a good value	.485	.712

Abbildung 8: Diese rotierte Komponentenmatrix veranschaulicht die Beziehung zwischen den Faktoren und den Umfragnaussagen. Durch die Analyse der den einzelnen Faktoren zugeordneten Aussagen ergibt sich, dass Faktor 1 die wahrgenommene Qualität misst, während Faktor 2 den wahrgenommenen Wert bzw. das wahrgenommene Preis-Leistungs-Verhältnis misst.

Faktoren sind definitionsgemäß valide, da alle zugeordneten Fragen bzw. Aussagen Aspekte desselben Konzepts messen. Es ist jedoch weiterhin wichtig, die Reliabilität (Zuverlässigkeit) des Faktors eindeutig festzustellen.

An dieser Stelle in der Analyse können Sie aussagekräftige Faktorbeschriftungen erstellen, die in Berichten und nachfolgenden Analysen verwendet werden sollen. In diesem Fall haben wir Faktor 1 als „Quality“ (Qualität) bezeichnet, da er verschiedene Aspekte der Qualität der Serviceleistungen und Einrichtungen des Hotels misst. Faktor 2 wurde mit „Value“ (Preis-Leistungs-Verhältnis) bezeichnet, da er die Zufriedenheit mit dem Zimmerpreis und den wahrgenommenen Wert der Serviceleistungen misst.

Die einzelnen Antworten auf die verschiedenen Aussagen sind zwar weiterhin nützlich, doch wir haben nun zwei Faktoren, die stichhaltigere Maßzahlen für die Qualität und das Preis-Leistungs-Verhältnis der Hotelkette bieten, als jede der Einzelaussagen für sich allein genommen. Dies ist der wichtigste Vorteil, den die Faktoranalyse bietet.

Um Faktoren in zukünftigen Analysen zu verwenden, erstellen wir eine kombinierte Maßzahl für die Fragen bzw. Aussagen, die den einzelnen Faktoren zugeordnet sind. Dies kann auf zwei verschiedene Weisen erreicht werden:

- Automatische Erstellung von Faktorwerten mit der Faktor-Prozedur und standardisierten Z-Scores
- Berechnung einer neuen Variablen durch Addieren der Rohantworten für die den einzelnen Faktoren zugeordneten Aussagen und anschließendes Dividieren durch die Anzahl der Fragen zur Berechnung eines Mittelwerts (mittlerer Score)

Verwendung eines Faktors in der Analyse

Doch was können Sie mit einem Faktor tun, nachdem Sie festgestellt haben, dass er valide und zuverlässig ist? Die Antwort ist einfach: Behandeln Sie den Faktor so, wie Sie eine neue Variable behandeln würden, und verwenden Sie ihn wie jede andere Frage, die Sie für vertrauenswürdig halten. Die Beziehungen sind bei Faktoren normalerweise klarer und eindeutiger als bei einzelnen Fragen oder Aussagen.

Zur Verdeutlichung untersuchen wir die Beziehung zwischen der Qualitätsbeurteilung und der Häufigkeit der Übernachtungen in BSI-Hotels. Zunächst verwenden wir die Einzelaussage zur Qualität (Abbildung 10). Anschließend führen wir dieselbe Analyse mit dem Faktor „Quality“ (Qualität) durch (Abbildung 11).

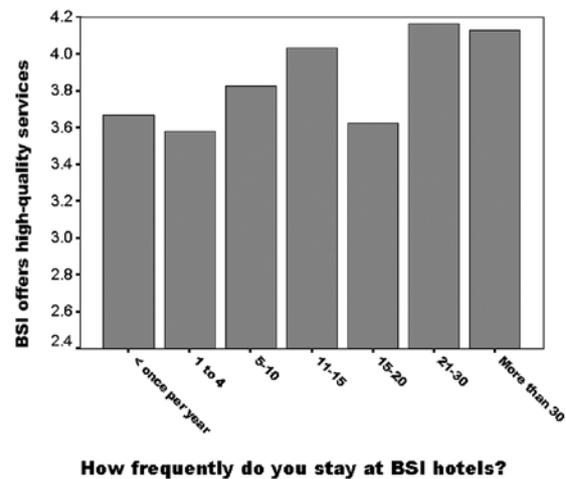


Abbildung 10: Dieses Diagramm zeigt die Ergebnisse der Analyse unter Verwendung der Einzelaussage zur Qualität. Die Ergebnisse besagen, dass Kunden, die häufiger in BSI-Hotels übernachten, mit etwas höherer Wahrscheinlichkeit zustimmen, dass die Kette hochwertige Serviceleistungen bietet.

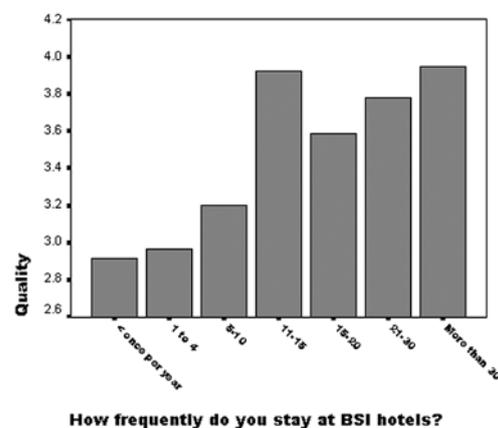


Abbildung 11: Dieses Diagramm zeigt die Ergebnisse der Analyse unter Verwendung des Faktors „Quality“ (Qualität). Die Ergebnisse weisen deutlicher darauf hin, dass Kunden, die häufiger in BSI-Hotels übernachten, mit größerer Wahrscheinlichkeit die Qualität als höher wahrnehmen und dass Qualität offenbar eine wichtige Komponente der Kundentreue ist.

Beide Diagramme zeigen eine positive Beziehung zwischen der wahrgenommenen Qualität und der Übernachtungshäufigkeit. Beachten Sie jedoch, dass die Beziehung für den Faktor „Quality“ (Qualität) wesentlich stärker ist als für die Einzelaussage zur Qualität. Die Qualität eines Hotels setzt sich aus mehreren Serviceaspekten zusammen und Einzelfragen sind definitionsgemäß nicht gut dafür geeignet, alle Aspekte eines allgemeinen Konzepts zu erfassen. Dieses Beispiel ist eine direkte Veranschaulichung der Vorteile, die die kombinierte Verwendung von Faktor- und Reliabilitätsanalyse bietet.

Mehrfachregression ist die beliebteste Methode zur Untersuchung der Beziehung zwischen einer Ergebnisvariablen und mehreren Prädiktorvariablen.

Vorhersagen mit Regression

Mehrfachregression, eine Technik vom Typ „Allgemeines lineares Modell“, ist die beliebteste Methode zur Untersuchung der Beziehung zwischen einer Ergebnisvariablen und mehreren Prädiktorvariablen (unabhängigen Variablen). Sie wird häufig bei Umfragedaten verwendet, weil sie die Zusammenfassung mehrerer Variablen zu einer einzigen Vorhersagegleichung ermöglicht. Darüber hinaus erleichtert die Mehrfachregression die Ermittlung der Rolle, die die einzelnen Variablen bei der Vorhersage des Ergebnisses spielen, sie stellt ein Maß für die Gesamterklärungskraft des Modells (R^2) bereit und sie bietet eine Schätzung darüber, ob eine Variable ein statistisch signifikanter Prädiktor ist oder nicht.

Die Mehrfachregression wird häufig als lineare Regression bezeichnet, da von einer linearen, also geradlinigen, Beziehung zwischen Prädiktoren und Ergebnis ausgegangen wird. Die Beziehungen zwischen den Variablen sind möglicherweise nicht immer linear, für die Erstellung eines brauchbaren Modells geht man jedoch am besten davon aus, dass sie es sind. Wie bei der Faktoranalyse funktioniert auch die Mehrfachregression am besten mit intervallskalierten Variablen, Sie können jedoch auch die typischen Umfrage-Antwortskalen verwenden.

Im Allgemeinen sollten bei der Regressionsanalyse nur Variablen verwendet werden, die gute Prädiktoren sein könnten, oder Variablen, die Sie aufnehmen möchten, weil praktische Gründe (beispielsweise wenn die Variable „Kundentyp“ für Ihr Unternehmen relevant ist) oder theoretische Gründe (frühere Arbeiten legen nahe, dass das Geschlecht der Kunden ein wichtiger Prädiktor ist) vorliegen. Es können zwar Dutzende Prädiktorvariablen in eine Regressionsgleichung aufgenommen werden, es ist jedoch sinnvoll, bei der Auswahl etwas gezielter vorzugehen.

Die ersten Schritte bei der Regressionsanalyse bestehen darin, die Prädiktoren und die Variable bzw. Frage, für die eine Vorhersage erstellt werden soll, anzugeben. In diesem Fall wenden wir Regressionstechniken auf die Daten der hypothetischen Hotelkette BSI an, um die Gesamtzufriedenheit der Kunden vorherzusagen.

Wir verwenden neun Prädiktoren, darunter den vorher ermittelten Faktor „Qualität“, um zu demonstrieren, wie Regressions- und Faktoranalyse miteinander verknüpft werden können.

Die ersten von Statistics erstellten Ergebnisse sind in den beiden Tabellen in Abbildung 12 zu sehen. Der R²-Wert beträgt 0,583. Diese neun Prädiktoren erklären also mehr als die Hälfte der Variation in den Antworten zur Kundenzufriedenheit. Die zweite Tabelle zeigt, dass die ausgewählte Menge an Prädiktoren auf dem Signifikanzniveau 0,01 statistisch signifikant für der Vorhersage der Kundenzufriedenheit ist (da der Wert „Sig.“ darunter liegt). Die Wahrscheinlichkeit dafür, dass sich unsere Regressionsergebnisse zufällig ergeben haben, beträgt also weniger als ein Hundertstel.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.768	.590	.583	.654

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	335.453	9	37.273	87.177	.000
	Residual	233.016	545	.428		
	Total	568.468	554			

Abbildung 12: Diese Modellzusammenfassungstabelle und die ANOVA-Tabelle zeigen die Ergebnisse der Regressionsanalyse für die Vorhersage der Gesamtzufriedenheit der Kunden.

Der nächste Schritt besteht darin zu ermitteln, welche der Fragen signifikante Prädiktoren im multivariaten Modell sind (Abbildung 13). Dazu überprüfen wir die Spalte „Sig.“ auf Werte unter ca. 0,05. Sieben der neun Fragen erreichen diesen Standard (eine nur knapp). Dies sind nützliche Informationen, denn sie sagen uns, welche Variablen signifikant sind und welche nicht. Wir sehen, dass die Zufriedenheit mit dem Zimmerpreis und den Einrichtungen bei der Vorhersage der Kundenzufriedenheit nicht wichtig ist.

Anschließend verwenden wir die nichtstandardisierten Koeffizienten in Spalte B, um den Effekt der einzelnen Prädiktoren zu bewerten. Beispielsweise bedeutet der Wert 0,468 in Spalte „B“ für „BSI services are a good value“ (BSI bietet bei seinen Services ein gutes Preis-Leistungsverhältnis), dass bei jedem Anstieg um eine Einheit bei der Antwort auf diese Frage die Zufriedenheit um 0,468 Einheiten steigt. Anders ausgedrückt: Wenn wir Kunden, die zustimmen (Score = 4), dass BSI bei seinen Services ein gutes Preis-Leistungsverhältnis bietet, mit Kunden vergleichen, die dieser Aussage voll und ganz zustimmen (Score = 5), sagen wir voraus, dass die Gesamtzufriedenheit bei letzteren um ca. 0,47 Einheiten höher liegt. Diese Beziehung kontrolliert die anderen Variablen im Modell. Mit anderen Worten: Wir können diese absolute Aussage treffen, da die Mehrfachregression automatisch alle Variablen im Modell kontrolliert. Durch die Verwendung der B-Koeffizienten für alle signifikanten Prädiktoren können wir eine Vorhersagegleichung für die Gesamtzufriedenheit der Kunden erstellen.

Beachten Sie, dass der B-Koeffizient für die Übernachtungshäufigkeit negativ ist. Das bedeutet, dass Kunden, die häufiger in BSI-Hotels übernachten, weniger zufrieden sind als unregelmäßige Kunden, nachdem andere Faktoren berücksichtigt wurden. Dies könnte eine wichtige Erkenntnis sein, der BSI-Analysten und die Geschäftsleitung nachgehen sollten.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.054	.198		.275	.783
BSI services are a good value	.468	.041	.521	11.444	.000
BSI facilities are kept up to date	.034	.033	.033	1.034	.301
Am satisfd with information received on services	.062	.032	.062	1.955	.051
BSI rooms are priced right	.021	.040	.022	.530	.597
Usefulness of internet services	-.051	.019	-.079	-2.642	.008
How frequently do stay at BSI hotels?	-.070	.018	-.112	-3.861	.000
GOVT	.156	.078	.059	2.010	.045
CORP	.146	.069	.066	2.116	.035
Quality	.520	.052	.351	9.934	.000

a. Dependent Variable: I am a satisfied BSI customer

Abbildung 13: Diese Regressionsanalysetabelle zeigt die signifikanten Prädiktoren für die Kundenzufriedenheit auf.

Der nächste Schritt besteht in der Ermittlung der relativen Wichtigkeit der Variablen mithilfe des Beta-Koeffizienten aus der Tabelle in Abbildung 14. Beta-Werte, die von -1 bis 1 reichen, sind wichtig, wenn für die Antworten unterschiedliche Antwortskalen gelten. Je höher der absolute Wert von Beta, desto wichtiger ist die Variable bei der Vorhersage der Kundenzufriedenheit. In diesem Fall sehen wir, dass die wichtigsten Prädiktoren der wahrgenommene Wert der Serviceleistungen sowie der Faktor „Qualität“ aus unserem Beispiel zur Faktoranalyse sind.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.054	.198		.275	.783
BSI services are a good value	.468	.041	.521	11.444	.000
BSI facilities are kept up to date	.034	.033	.033	1.034	.301
Am satisfd with information received on services	.062	.032	.062	1.955	.051
BSI rooms are priced right	.021	.040	.022	.530	.597
Usefulness of internet services	-.051	.019	-.079	-2.642	.008
How frequently do stay at BSI hotels?	-.070	.018	-.112	-3.861	.000
GOVT	.156	.078	.059	2.010	.045
CORP	.146	.069	.066	2.116	.035
Quality	.520	.052	.351	9.934	.000

a. Dependent Variable: I am a satisfied BSI customer

Abbildung 14: In dieser Tabelle sind die Beta-Werte der beiden wichtigsten Prädiktoren hervorgehoben.

Auf der Grundlage dieser Ergebnisse könnte die BSI-Hotelkette eine effiziente und wirksame Strategie zur Verbesserung der Kundenzufriedenheit entwickeln, die sich nur auf die relevantesten Merkmale konzentriert. Außerdem könnte die Kette diese Informationen verwenden, um Vorhersagen über unzufriedene Kunden zu treffen und zukünftige Marketingaktionen entsprechend auszurichten.

Beim BSI-Beispiel können wir mithilfe der Regression erkennen, dass die Kundenzufriedenheit in engerem Bezug zum wahrgenommenen Wert der Serviceleistungen steht als zum Zimmerpreis. Obwohl „BSI bietet bei seinen Services ein gutes Preis-Leistungsverhältnis“ der beste Prädiktor überhaupt ist, ist „Der Preis der Zimmer bei BSI ist gut“ nicht einmal ein signifikanter Prädiktor. Dies überrascht, da man annehmen könnte, dass der Preis sich auf die Zufriedenheit auswirkt. Diese Entdeckung legt nahe, dass der Vergleich zwischen Preis und wahrgenommenem Wert wichtiger für die Kunden ist als der Preis allein. Durch die Regressionsanalyse erhalten wir wichtige Einblicke in die Beziehung zwischen diesen beiden Faktoren und die tatsächlichen Haltungen der BSI-Kunden.

Zusammenfassung

Wie in den Beispielen in diesem Dokument dargelegt, können Sie mit fortschrittlichen Methoden zur Umfrageanalyse wertvolle Detailinformationen zu den Haltungen und Verhaltensweisen von Umfrageteilnehmern extrahieren.

IBM SPSS Statistics bietet leistungsfähige Techniken, wie die hier erörterten Cluster-, Faktor-, Reliabilitäts- und Regressionsmethoden, mit denen Sie Beziehungen aufdecken können, die Ihnen mit einfacherer Software und grundlegenden Methoden möglicherweise entgehen. Für die Präsentation Ihrer Ergebnisse stellt Statistics die professionellen Grafik- und Berichtsfunktionen bereit, die Sie benötigen, um komplexe Ideen effektiv zu vermitteln.

Die in diesem Dokument erörterten fortschrittlichen Methoden können von jedem Umfrageforscher verwendet werden, da Statistics die Analysen intuitiv und unkompliziert macht. Je häufiger Sie die hier dargelegten Methoden einsetzen, desto leichter wird es Ihnen fallen und desto größere Einblicke können Sie aus Ihren Umfragedaten gewinnen. Außerdem können diese fortschrittlichen Methoden Ihrem Unternehmen einen Wettbewerbsvorteil verschaffen und Sie in die Lage versetzen, zeitnah und effektiv fundiertere Entscheidungen zu treffen.

Informationen zu IBM Business Analytics

Die Software IBM Business Analytics liefert umfassende, einheitliche und korrekte Informationen, denen Entscheidungsträger zum Verbessern der Unternehmensleistung vertrauen. Ein umfassendes Portfolio aus Geschäftsvorteilen, fortgeschrittener Analytik, finanziellen Vorteilen und Strategiemanagement sowie Analyseanwendungen bietet Ihnen sofort klare und umsetzbare Einblicke in die aktuelle Leistung und gibt Ihnen die Möglichkeit, zukünftige Ergebnisse vorherzusagen.

Als Teil dieses Portfolios unterstützt IBM SPSS Predictive Analytics Software Organisationen, zukünftige Ereignisse vorherzusagen und proaktiv auf Basis dieser Erkenntnisse zu handeln, um bessere Geschäftsergebnisse zu erzielen. Kunden aus den Bereichen Wirtschaft, öffentlicher Verwaltung und Lehre verlassen sich weltweit auf IBM SPSS Technologie als Wettbewerbsvorteil zur Kundengewinnung, -bindung und Erhöhung der Kundenumsätze bei gleichzeitiger Betrugsreduzierung und Risikominimierung. Durch die Integration von IBM SPSS Software in ihre täglichen Prozesse werden Organisationen zur Predictive Enterprise – sie sind dadurch in der Lage Entscheidungen zu treffen und zu automatisieren, um die Geschäftsziele zu erreichen und einen messbaren Wettbewerbsvorteil zu gewinnen. Für mehr Informationen besuchen Sie bitte www.ibm.com/spss/de.



© Copyright IBM Corporation 2010

IBM Corporation
Route 100
Somers, NY 10589

Eingeschränkte Rechte für Mitarbeiter der US-Regierung – Benutzung, Duplizierung und Veröffentlichung beschränkt durch GSA ADP Schedule-Vertrag mit IBM Corp.

Hergestellt in den USA
Mai 2010
Alle Rechte vorbehalten

IBM, das IBM Logo, ibm.com, WebSphere, InfoSphere und Cognos sind Marken oder eingetragene Marken der International Business Machines Corporation in den USA und/oder anderen Ländern. Wenn diese oder andere eingetragene Markenbegriffe von IBM mit einem Markenzeichen (® oder TM) gekennzeichnet sind, wenn Sie zum ersten Mal in diesen Informationen vorkommen, weist dies darauf hin, dass es sich zum Zeitpunkt der Veröffentlichung dieser Informationen um in den USA eingetragene Marken oder Marken nach Gewohnheitsrecht von IBM handelt. Solche Marken können auch in anderen Ländern eingetragene Marken oder Marken nach Gewohnheitsrecht sein. Eine aktuelle Liste der Marken von IBM finden Sie im Internet unter „Copyright and trademark information“ (www.ibm.com/legal/copytrade.shtml).

SPSS ist eine Marke von SPSS, Inc., an IBM Company, die in vielen Gerichtsbezirken weltweit eingetragen ist.

Weitere Unternehmens-, Produkt- und Servicenamen können Marken anderer Hersteller sein.



Bereiten Sie bitte auf